

**AD-A284 762**



**FINAL REPORT:**

**FUSION OF MULTIPLE SENSING  
MODALITIES FOR MACHINE VISION**

**J. K. Aggarwal, Principal Investigator**

**February 1, 1991 - May 31, 1994**

**U. S. ARMY RESEARCH OFFICE**

**Contract DAAL03-91-G-0050**



**Computer and Vision Research Center**

**The University of Texas at Austin**

**Austin, Texas 78712**

**512/471-3259**

**APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED**

*3918*

**94-30696**



**94 9 22 105**

THE VIEWS, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHORS AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 5/31/94	3. REPORT TYPE AND DATES COVERED Final, 2/1/91 - 5/31/94	
4. TITLE AND SUBTITLE Fusion of Multiple Sensing Modalities for Machine Vision			5. FUNDING NUMBERS Contract DAAL-03-91-G-0050  Proposal 28258-PH	
6. AUTHOR(S)  J. K. Aggarwal				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  The University of Texas at Austin Austin, TX 78712-1084			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  We report on a broad program of research in machine vision to develop an approach based upon synergistically combining diverse sensing modalities. The research projects fall into four general categories: Outdoor Scene Interpretation via the Fusion of Multiple Imaging Modalities; (2) Motion Computation and Object Recognition Using Range Images; (3) Structure and Identity Based on Color and Shape Information; and (4) Autonomous Navigation. Accomplishments include the development of the AIMS (automatic interpretation using multiple sensors) knowledge-based system to interpret registered laser radar and thermal images for the detection and recognition of man-made objects in outdoor rural scenes; the development of a new approach for the detection of large man-made objects using perceptual organization techniques; new algorithms for object recognition and motion estimation, including improved algorithms for using three-dimensional (range) images to compute structure and motion; a CAD-based object recognition system; a decision-theoretical algorithm to estimate 3D structures from extended sequences of 2D images taken by a moving camera; an algorithm for matching line segments based on perceptual grouping relaxation labeling; and the construction of an autonomous mobile robot, Robo-Tex, as a testbed for navigation algorithms.  DTIC QUALITY INSPECTED 3				
14. SUBJECT TERMS multisensor, position estimation, autonomous navigation, correspondence, segmentation, object recognition			15. NUMBER OF PAGES 40	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

# **FUSION OF MULTIPLE SENSING MODALITIES FOR MACHINE VISION**

## **FINAL REPORT**

**J. K. AGGARWAL, PRINCIPAL INVESTIGATOR**

**MAY 31, 1994**

**U. S. ARMY RESEARCH OFFICE**

**CONTRACT DAAL-03-91-G-0050**

**COMPUTER AND VISION RESEARCH CENTER  
THE UNIVERSITY OF TEXAS AT AUSTIN  
AUSTIN, TEXAS 78712-1084**

**APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED**

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification .....	
By .....	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

# **FINAL REPORT: FUSION OF MULTIPLE SENSING MODALITIES FOR MACHINE VISION**

## **SUMMARY**

The Computer and Vision Research Center at The University of Texas at Austin has undertaken a broad program of research in machine vision to develop an approach based upon synergistically combining diverse sensing modalities. The research projects funded under Contract DAAL-03-91-G-0050 fall into four general categories: (1) Outdoor Scene Interpretation via the Fusion of Multiple Imaging Modalities; (2) Motion Computation and Object Recognition Using Range Images; (3) Structure and Identity Based on Color and Shape Information; and (4) Autonomous Navigation.

Some of the highlights of our accomplishments include the development of the AIMS (automatic interpretation using multiple sensors) knowledge-based system to interpret registered laser radar and thermal images for the detection and recognition of man-made objects in outdoor rural scenes, including a new algorithm for integration of region and edge information without the intervention of high-level knowledge, and the development of a new approach for the detection of large man-made objects using perceptual organization techniques. We have developed a number of new algorithms for object recognition and motion estimation, including improved algorithms for using three-dimensional (range) images to compute structure and motion; a CAD-based object recognition system which uses a three-dimensional CAD model of an object to locate the object in a cluttered scene; a decision-theoretical algorithm to estimate 3D structures from extended sequences of 2D images taken by a moving camera; and an algorithm for matching line segments based on perceptual grouping relaxation labeling. Finally, a significant body of work has been accomplished in the area of autonomous navigation with the construction of an autonomous mobile robot, Robo-Tex, as a testbed for navigation algorithms, as well as a number of projects on position estimation and calibration techniques.

Significant research findings from these research projects have been presented at national and international conferences and published in referred journals. A total of 13 graduate and 4 undergraduate students were supported under this contract, and 6 Ph.D. and 1 M.S. degrees were completed during the contract term.

**FINAL REPORT:**  
**FUSION OF MULTIPLE SENSING MODALITIES**  
**FOR MACHINE VISION**

**TABLE OF CONTENTS**

	<b>Page</b>
<b>STATEMENT OF THE PROBLEM STUDIED</b>	<b>1</b>
<b>SUMMARY OF IMPORTANT RESULTS</b>	<b>3</b>
1. Outdoor Scene Interpretation via the Fusion of Multiple Imaging Modalities	5
a. Automatic Interpretation Using Multiple Sensors	5
b. Locating Man-made Objects in Outdoor Scenes	7
c. Modeling Non-Homogeneous 3-D Objects for Thermal and Visual Image Synthesis	8
d. Object Recognition Using ART-2 Artificial Neural Network	9
2. Motion Computation and Object Recognition Using Range Images	11
a. Surface Correspondence and Motion Computation from a Sequence of Range Images	11
b. Segmentation of 3D Range Images Using Pyramidal Data Structures	12
c. The Convergence of Fuzzy Pyramid Algorithms	12
d. CAD-Based Object Recognition	12
3. Structure and Identity Based on Color and Shape Information	15
a. Integration of Image Segmentations Maps Using Region and Edge Information	15
b. Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot	16
c. 3D Structure Reconstruction from an Ego Motion Sequence Using Statistical Estimation and Detection Theory	17

d.	Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images	18
e.	Stereo Image Interpretation in the Presence of Narrow Occluding Objects	18
f.	Analysis of Video Images Using Point and Line Correspondences	19
4.	Autonomous Navigation	21
a.	Construction of an Autonomous Mobile Robot, Robo-Tex	21
b.	Position Estimation Techniques for an Autonomous Mobile Robot	23
c.	Calibrating a Mobile Camera's Parameters	25
d.	Calibration Procedure for a Fish-Eye Lens (Super-Wide-Angle Lens) Camera	25
	CONCLUSIONS	27
	BIBLIOGRAPHIC REFERENCES	29
	PUBLICATIONS AND TECHNICAL REPORTS	33
	PARTICIPATING SCIENTIFIC PERSONNEL	39
	LIST OF INVENTIONS	39
	LIST OF FIGURES AND TABLES	40

## STATEMENT OF THE PROBLEM STUDIED.

Machine vision—the automated interpretation of sequences of images to detect and recognize, locate and track objects—has many important applications in the peace-keeping activities of the Department of Defense, including automated surveillance and monitoring, autonomous navigation for smart weapons, and industrial robotics. For machine vision systems to be truly useful, information must be extracted quickly from digitized images. Extraction is a complex task: large amounts of data must be processed, noise is present in the images, information may be incomplete, or the models of the scene and sensors may be inadequate. To establish useful and practical methods for machine perception of targets and guidance of payloads, a broad program of research in machine vision is needed.

Previous research in machine perception focused mainly on the use of a single sensing modality, for example, a video or an infrared camera. However, most single sensor systems work only in highly constrained environments and require massive computational resources. These limitations can be overcome by using multiple sensing modalities and developing “intelligent” algorithms to effectively combine these modalities.

In a broad program of research in machine vision at The University of Texas at Austin, we have developed an approach in which diverse sensing modalities are synergistically combined. The synergistic fusion of information from multiple sensors can discern additional features that provide better discrimination than can be obtained by processing the sensor inputs separately. Our work has focused upon building physical models of the scene in order to relate the signals obtained from different modalities to the various parameters of objects in the scene. Based on these estimated parameters, we can identify and evaluate intrinsic properties of the objects in order to interpret the scene.

Four broad areas of inquiry were pursued under this contract, as outlined below:

- (1) Outdoor Scene Interpretation via the Fusion of Multiple Imaging Modalities
- (2) Motion Computation and Object Recognition Using Range Images
- (3) Structure and Identity Based on Color and Shape Information
- (4) Autonomous Navigation

Significant results from the projects in these areas of inquiry are summarized below.

## **SUMMARY OF IMPORTANT RESULTS.**

The research projects carried out under this contract have led to significant progress in the study of some of the most difficult problems in computer vision, including multisensor systems for outdoor scene interpretation, algorithms for motion computation and object recognition in cluttered environments using range images, techniques for computing structure and identity based on color and shape data, and the development of a mobile robot, Robot-Tex, as a testbed for autonomous navigation algorithms.

## 1. Outdoor Scene Interpretation via the Fusion of Multiple Imaging Modalities.

### a. Automatic Interpretation Using Multiple Sensors.

The AIMS (automatic interpretation using multiple sensors) knowledge-based system was developed to interpret registered laser radar and thermal images to detect and recognize man-made objects, such as armored personnel carriers and trucks, in outdoor rural scenes with a background of vegetation, ground, and sky. [1-6]. The system applies the multisensor fusion approach to multiple ladar modalities to improve both segmentation and interpretation. An early version of the system, using laser radar and thermal images, was presented at the 7th IEEE Conference on Artificial Intelligence Applications (1991), where it received the conference's outstanding paper award [1]. The system was later expanded to four sensing modalities, range, intensity, velocity, and thermal) to improve image segmentation and interpretation.

The interpretation system developed [2] is not limited to the domain of object recognition, and could also be used for robot navigation and obstacle avoidance. The design of the rule bases is modularized to permit future expansion of the system to incorporate additional sensing modalities. The system applies the *multisensor fusion* approach to multiple ladar modalities to improve both segmentation (pixel-level sensor fusion) and interpretation (object-level sensor fusion). This approach offers the dual advantages of (1) the ability to work under less than optimal imaging environments (rain, night, etc.) and (2) the ability to detect objects and estimate their attributes with better precision.

The use of different sensors provides not only different types of information, but also multiple observations of the same information through different channels. The knowledge-based interpretation system is constructed using KEE and Lisp. Low-level attributes of image segments (regions) are computed by the segmentation modules and then converted to the KEE format. The interpretation system applies forward chaining in a bottom-up fashion to derive object-level interpretation from input generated by low-level processing and segmentation modules. The interpretation modules detect man-made objects from the background using low-level attributes. Segments are grouped into objects, which are then classified into predefined categories (vehicles, ground, etc.). The efficiency of the interpretation system is enhanced by transferring nonsymbolic processing tasks to a concurrent service manager (program).

The AIMS system is based upon a new integration algorithm that integrates multiple region segmentation maps and edge maps [5]. It operates independently of image sources and specific region-segmentation or edge-detection techniques. User-specified weights and the arbitrary mixing of region/edge maps are allowed. The integration algorithm enables multiple edge detection/region segmentation modules to work in parallel as front ends. The solution procedure consists of three steps. A maximum likelihood estimator provides initial solutions to the positions of edge pixels from various inputs. An iterative procedure using only local information (without edge tracing) then minimizes the contour curvature. Finally, regions are merged to guarantee that each region is large and compact. The channel-resolution width controls the spatial scope of the initial estimation and contour smoothing to facilitate multiscale processing. Experimental results are demonstrated using data from different types of sensors and processing techniques. The results show an improvement over individual inputs and a strong resemblance to human-generated segmentation.

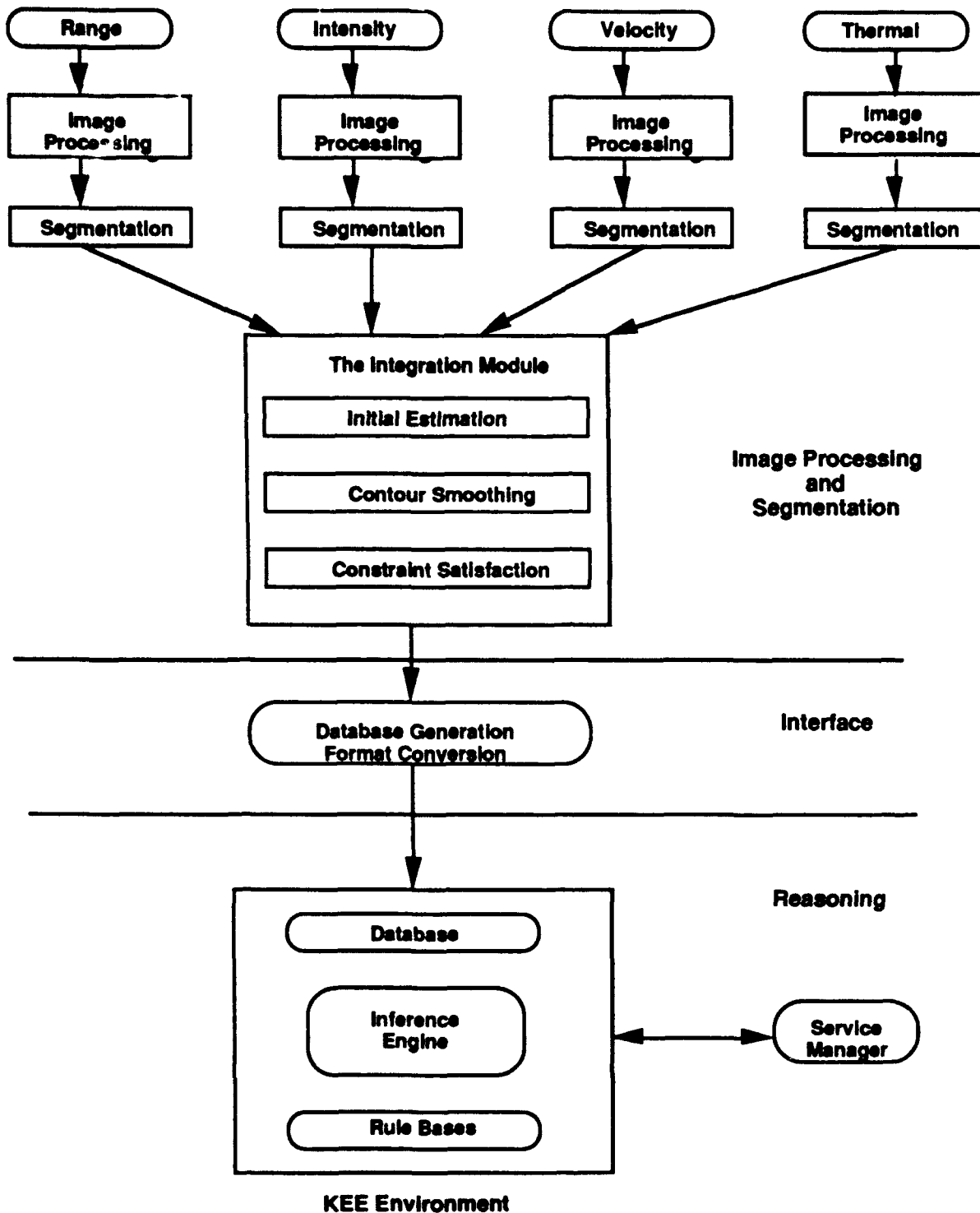


Figure 1. The overall structure of AIMS.

The AIMS system has been ported to an AT&T Pixel Machine to study the behavior of a rule-based system for image understanding in a multiprocessor environment and to study the hardware and software requirements for such an implementation [6]. The AT&T Pixel Machine employs a distributed memory computer architecture with message passing. Past research efforts

in parallel processing of rule-based systems used the fine-grain *Rete* algorithm on less powerful processors. However, with the increased computing power and decreased cost of new multiprocessor hardware, new software strategies are needed. The system generates a rule firing rate of 700+ rules/second and has a linear speedup with respect to the number of processors, on a platform designed for graphics rendering. The experiment shows that given strong data locality, a distributed memory architecture can support complex software systems, including rule based systems, without using fine-grained parallelism.

b. Locating Man-made Objects in Outdoor Scenes.

A new approach for the detection of large man-made objects was developed using perceptual organization techniques to group low level features in a scene to determine a region of interest that is most likely to contain man-made objects [7-9]. In this study, the man-made objects may be unspecified and the appearance of the objects is unpredictable. The approach applies the principles of perceptual organization and makes use of prominent features that distinguish man-made objects from natural objects. Using computer vision techniques such as feature extraction, primitive structure formation, and segmentation, we hierarchically group low-level image features into a region-of-interest, an area which is deemed mostly likely to enclose man-made objects or a substantial part of a man-made object. A paper based upon this work was published in *Pattern Recognition*, and received the Honorable Mention of the Pattern Recognition Society Award for Outstanding Contribution, November 1993.

The goal of our research is to detect man-made objects from images of natural scenes. Since the objects are not particularly specified, features must be found that distinguish man-made objects from natural objects in an image. Two of the most prominent characteristics of manmade objects are the apparent regularity and relationship of their components. Most man-made objects have linear structures or linear boundaries that form certain regular patterns, such as rectangles, parallels and polygons. These regular patterns are usually related to each other and form the man-made objects. After line detection, much of such regularity and relationship remains apparent. To detect the man made objects, we must extract the geometric structures that exhibit regularity and relationship from the image. Hence, the framework of our approach includes three phases: (1) extracting image features, (2) finding regularities and relationships among these features, and (3) identifying the region occupied by the related regular structures.

The first level of grouping extracts image features. We currently consider two kinds of features: *linear structures* (LS) and *coterminations* (CT). A linear structure in an image is a representation of a set of approximately collinear line segments which are close and likely to come from the same linear structure in the scene. Extracting LS reflects the proximity, collinearity, and continuation properties of perceptual grouping. A cotermination is a set of lines terminating at a common point or a small common region. The cotermination is an important relation. Cotermination is a non-accidental relationship and, hence, reflects significant structural information. It is also view invariant in a wide range of viewpoints and can be used for 3D inference. The CT are represented by a graph called the *CT graph*.

The second level grouping process organizes features that exhibit regularity and relationship into larger structures called *primitive structures* (PS). We consider two kinds of PS: *parallel PS* and *polygon PS*. A parallel PS is a set of parallel lines satisfying certain conditions. A polygon PS is a closed figure that consists of line segments and satisfies certain criteria. Parallel and polygon PS are higher level structures than lines and coterminations. The PS are represented by a graph called the *PS graph* that describes the spatial relationships among the PS. Such a graph facilitates higher level processing. Using the PS graph, the third phase of the framework groups spatially closed PS, eliminates the isolated ones, and segments the image into regions occupied by the grouped PS and a background. The largest region of the grouped PS is then evaluated based on the area of the region and the statistics of the PS. If this region is

determined to be significant, it is most likely to enclose man-made objects or a substantial part of the man-made objects, and thus is considered to be the *region of interest* (ROI). Figure 2 illustrates the overall data representation, relationship, and flow in this framework.

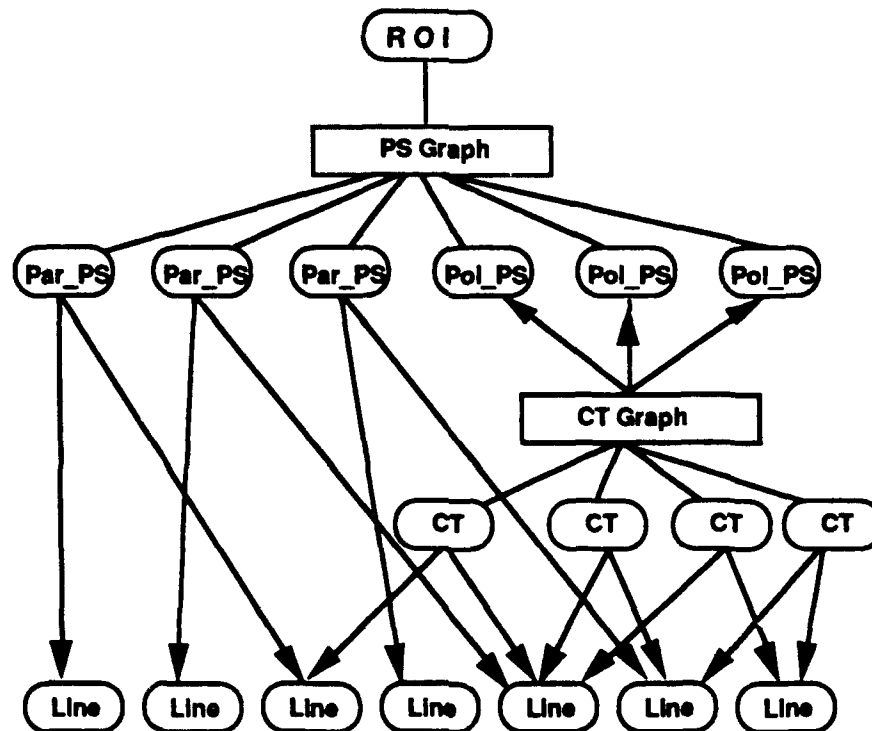


Figure 2. The data representation and data relationship in the framework.

The study used monochrome images containing man-made objects such as bridges and electric transmission towers in complex backgrounds. By locating the region of interest in the image, the search space is substantially reduced from the whole image to that region. This technique could be useful for screening a large number of images for automatic object recognition or for a human-machine system. For an automatic system, when specific object classes are given and models are established, the primitive structures composing the region of interest can be matched to object models rather than to individual features. This will considerably reduce the search space for matching, since more constraints are applied. For a human-machine system, the region of interest can be used as a focus-of-attention for human expertise to further examine the image.

c. Modeling Non-Homogeneous 3-D Objects for Thermal and Visual Image Synthesis.

An approach to the integrated modeling of 3-dimensional objects was developed that supports the synthesis of visual and thermal images under different viewing, ambient, and internal conditions [10]. Object modeling is accomplished using the volume surface octree, a representation which is well suited for thermal modeling of complex objects with non-homogeneities and heat generation. An improved technique for constructing the volume surface octree increases storage efficiency without degrading the quality of the image synthesis. A technique is used to incorporate non-homogeneities and heat generation using octree intersection. A computationally efficient, implicit fine difference method is used to simulate heat flow of objects with a large number of octree nodes and with non-homogeneities. The model is designed to be used in

a multisensor vision system that would use the images and features predicted by this model in a hypothesize-and-verify scheme.

d. Object Recognition Using ART-2 Artificial Neural Network.

The Hybrid Architecture for Man Made Object Recognition (HAMMER) is an object recognition system that uses the Adaptive Resonance Theory (ART)-based ART-2 artificial neural network [11]. ART networks are one of the most promising neural network architectures for image recognition applications. The goal of this object recognition system is to recognize and classify man-made moving objects from a sequence of 2D images. The objects appear in natural scenes containing trees, buildings, landscapes, etc. As shown in Figure 3, the HAMMER system architecture incorporates a preprocessing module that extracts invariant features from the input image, which are then used by the neural network for object classification and recognition. The preprocessing module consists of two stages. In the first stage, the object (image figure) to be recognized is segmented from the image background. This is accomplished by determining the Region of Interest (ROI) in the image. In the second stage, a transformation is applied to the image, whose output spectra are invariant with image transformations such as 2D rotation, scaling and translation. The features extracted from this final preprocessing stage are then input to ART-2 for unsupervised classification. Based on the output of the ART-2 module, the DECISION LAYER then labels (i.e., names) the input object. The HAMMER system is implemented in C++ on a PC 486 and was tested using images of different types of military vehicles in natural surroundings. The HAMMER system correctly classified about 90% of the input objects. Overall, this system is capable of performing successfully in a complex, unknown environment.

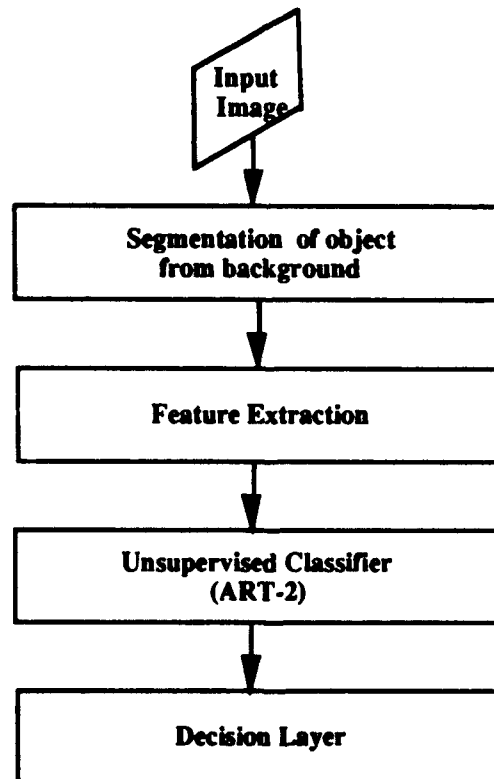


Figure 3. The HAMMER system

## 2. Motion Computation and Object Recognition Using Range Images.

Computer vision techniques play a significant role in a number of fields of engineering such as robotics and manufacturing. In most applications it is important to sense the environment and estimate the relative motion between the sensor and the objects in the environment. For example, when a robot moves in a work space it is important for the robot to know its motion relative to the other objects in the work space.

Traditionally, many vision systems have used a video camera as the primary sensor. However, video images provide no depth information about the scene. Human beings have remarkable vision systems that can extract 3-D information from 2-D images, but, at the present state of technology, it is impossible for a computer to achieve the same level of performance. However, this shortcoming can be overcome by using range image sensors in the computer vision system. A range sensor directly obtains the 3-D depth information of the scene. This information (called a 3-D range image) now can be used by the computer to reason about the environment and accomplish complex tasks.

### a. Surface Correspondence and Motion Computation from a Sequence of Range Images.

We have addressed the problem of determining the motion of the range sensor as the sensor moves relative to the objects in the scene [12-14]. The motion must be estimated from a 3-D range image sequence obtained from the range sensor. The key to solving this problem is identifying significant features in the images and using those features to compute the motion transformation. The range sensor senses the points on the object surface and generates a 3-D map of the scene. In such an imaging modality it is logical to use the object surfaces as the features to be used in the motion computation task. We have developed surface-based image processing techniques that are used in high level vision tasks.

The object surfaces are extracted from each 3-D range image. The next task is to track each surface segment over the image frames in the sequence. This is a fairly complex task, and to add to the complexity, the surfaces may become hidden by other surfaces as the sequence progresses, or new surfaces that were hidden earlier may appear in the images. The vision system developed at our research center handles such occurrences and reliably arrives at the solution.

The system uses the geometry and the topology of the scene to establish the correspondence between surfaces in different frames of the image sequence. This task is facilitated by the use of a representation scheme based on a hypergraph. After establishing correspondence between surface segments in the image frames, the next task is to compute the motion transformation. The computation of rigid motion transformation is a nonlinear problem to which the solution is not simple. If the surfaces are planar the motion computation task is significantly simplified, and the rotations and the translations can be estimated with reasonable certainty. However, in the case of the next order of surfaces, i.e., quadric surfaces, the motion computation equations become intractable. There is no guarantee that the solution obtained is the correct and optimal solution. We solve this problem by appealing to the geometry of the quadric surfaces. It has been known that every quadric has a unique linear feature. We extract this linear feature and use it to compute motion. Our approach uses object surfaces as the basis of all the tasks involved in motion computation. This results in reliable estimates of the motion transformation and the computation of correspondences. The procedure to compute motion using the linear features of quadric surfaces is novel and overcomes the problems faced by earlier approaches that

formulated complex non-linear optimization problems to compute the motion transformation. This approach can also be extended to other higher order surfaces.

b. Segmentation of 3D Range Images Using Pyramidal Data Structures.

The first step in most computer vision systems is to represent the acquired data using symbolic descriptions. Such descriptions are then used to perform higher level vision tasks, such as object recognition, motion estimation, or navigation. To obtain the symbolic descriptions, the input data must be partitioned, or segmented, into a set of primitives. The segmentation depends on the nature of the input data and on the primitives used by the high level tasks. We have addressed the problem of segmenting data that represent the 3D coordinates of each point in the scene (i.e., dense range images). Specifically, the problem is stated as: *Given a 3D range image of a scene containing multiple arbitrarily shaped objects, segment the scene into homogeneous surface patches.* We have proposed a new modular framework for this segmentation task [15]. The segmentation task is addressed within a framework of a vision system in which the output of the segmentation module is not the final objective. Instead, the segmentation procedure output must be capable of interpretation by the higher level modules. The high level vision tasks dictate to the segmentation module the criteria for uniformity of regions and the representation of the output.

The framework consists of two independent modules, the first of which performs low level segmentation and the second carries out the subsequent merging of regions. The segmentation is accomplished by an iterative pyramidal clustering scheme using zeroth and first-order local surface properties at each point of the scene. The segmentation is then refined in the second module using high order surface representations dictated by the upper level vision tasks.

The procedure has been applied successfully to many range images obtained from various institutions. This procedure offers a number of advantages over existing segmentation procedures. By using the modular framework, the low level segmentation process is independent of the surface type and description, and the high level process is independent of the local properties derived from the input data as well as from the method used to achieve the oversegmentation. Further, no restrictions are placed on the type or size of objects in the scene. Unlike most existing segmentation schemes, the procedure's dependency upon empirically determined thresholds is minimal. Finally, the pyramidal algorithms may be implemented in parallel.

c. The Convergence of Fuzzy Pyramid Algorithms.

Pyramid linking is an important technique for segmenting images and has many applications in image processing and computer vision. The algorithm is closely related to the ISODATA clustering algorithm and shares some of its properties. We have investigated this relationship and developed a proof of convergence for the pyramid linking algorithm [16]. The convergence of the "hard" pyramid linking algorithm has been shown in the past; however, there has been no proof of the convergence of "fuzzy" pyramid linking algorithms. The proof of convergence is based on Zangwill's theorem, which describes the convergence of an iterative algorithm in terms of a "descent function" of the algorithm. We show the existence of such a descent function on the pyramid algorithm, and demonstrate that all the conditions of Zangwill's theorem are met; hence, the algorithm converges.

d. CAD-Based Object Recognition.

We have addressed the problem of CAD-based object recognition, in which the objective is use a three-dimensional CAD model of an object to locate that object in a scene containing several overlapping objects, arbitrarily positioned and oriented [17-20]. A laser range

scanner is used to collect 3D data points from the scene. The collected data is segmented into surface patches, and the segments are used to calculate various 3D surface properties. CAD models are designed using commercially CADKEY and accessed via the industry standard IGES. The models are analyzed off-line to derive various geometric features, their relationships, and their attributes. A strategy for identifying each model is then automatically generated and stored. The strategy is applied at run-time to complete the task of object recognition. The goal of the generated strategy is to select the model's geometric features in the sequence which may best be suited to identify and locate the model in the scene. The generated strategy is guided by several factors, including the visibility, detectability, frequency of occurrence, and topology of the features.

This object recognition system significantly differs from previous systems, in that it uses a commercial 3D CAD system and IGES interface. Using CADKEY and IGES decreases the dependency of the vision system on any particular CAD modeler and increases its applicability. Moreover, the model description, derived automatically from IGES, is used to systematically derive a matching strategy from a geometric model. By using the recognition strategy, it is not necessary to consider all the possible matching combinations of sensory features and model features, increasing the efficiency of the system. Precompiling the recognition strategy also increases the vision system's run-time efficiency, since less time is spent on model analysis during task execution. Finally, the matching strategy is not significantly dependent on moment-based and boundary-based features, and, unlike many previous approaches to object recognition, the system does not require an unconditional one-to-one matching of sensory features and model features. Thus, the recognition system is not sensitive to the partial occlusion of objects, and the oversegmentation of surface patches is easily tolerated.

### 3. Structure and Identity Based on Color and Shape Information

#### a. Integration of Image Segmentations Maps Using Region and Edge Information.

As mentioned briefly above in Section 1(a), an algorithm has been developed for the AIMS interpretation system that integrates image segmentation maps using region and edge information without the intervention of high-level knowledge [5]. One problem encountered in using multiple sensing modalities is that when different segmentation techniques are applied to the images obtained from different modalities, different segmentation maps are generated. One has to resolve the differences between all such segmentation maps to benefit from the rich information provided by various sources. Information integration is a suitable approach to enhance system performance by verifying cues from one source to another. It is also necessary because of significant information loss during the image acquisition process. Information integration improves the signal-to-noise ratio, because information that is consistent among different sources is reinforced, while information that is contradicted is attenuated.

This algorithm integrates segmentations from different sensing modalities, segmentation techniques, and control parameters. It operates independently of the problem domains, the segmentation techniques, and any combination of edge and region maps. The basic task of the algorithm is estimation by generating a consensus of the true underlying segmentation from multiple observations. Further, the algorithm allows the flexibility of user-specified weights on different information sources, since they may not be equally reliable.

We assume that a true region contour map exists as the signal source, but that the signal is contaminated by noise during image acquisition, preprocessing, and segmentation. The objective of the integration module is to recover the original contour map from multiple contaminated copies using minimal knowledge about the signal and noise sources.

The work uses only the contour (position and length), the size, and the neighboring relationship attributes of regions. Other regional information is not used, because it would require the integration module to know about the construction of the front-end (region-growing-based) segmentation algorithms and about their operational results before a contour is generated. As shown in Figure 4, the integration procedure consists of three stages: (1) initial estimation--estimating the true region contours from given multiple observations, (2) contour smoothing--directly reducing curvature, and (3) constraint satisfaction--satisfying additional nonnegotiable constraints on the integration output according to the application context. The integration problem, therefore, is formulated as: *Given several sets of edge pixels and associated weights, solve for another set of edge pixels representing all input sets and exhibiting certain properties.* Our solution procedure decomposes the original formulation into three stages to pursue a suboptimal solution. The first stage is an estimator that generates an initial solution, while ensuring that the solution is a negotiated result from all the inputs and that the output is a sufficient representative of the input data. The second stage employs a potential-energy model to minimize the smoothness iteratively. The third stage checks the unnegotiable constraints and merges regions that violate such constraints. The work uses *maximum likelihood estimation* as the estimation strategy. *A priori* information is not considered, since it is usually unavailable in practical situations. The algorithm addresses the issue of figural continuity and incorporates this concern into the solution procedure. During the initial estimation process, continuity influences which data points are considered in the weighted average operation. After the initial estimation, the contour connectivity constraint enforces connections between pixels that are originally connected. The curve smoothing stage uses the *spring-node* model to refine node positions iteratively, subject to the *finite-drift* and *connected-neighbor* constraints. The algorithm treats juncture and nonjuncture pixels uniformly, and satisfies nonnegotiable constraints on region size and contour compactness if a region map is desired.

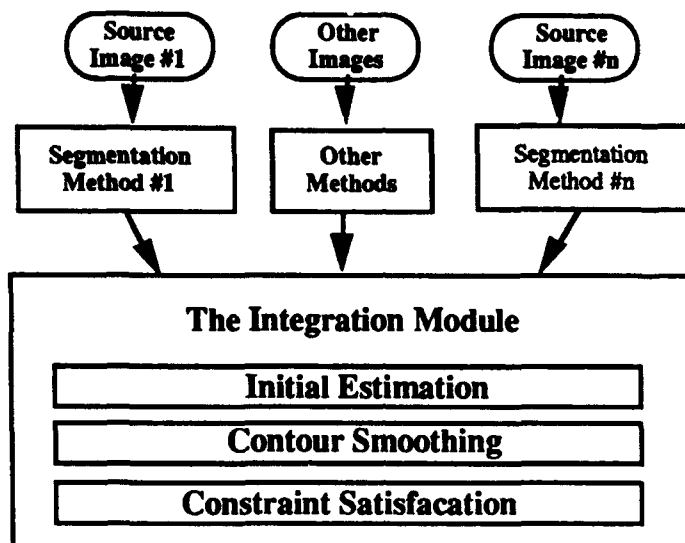


Figure 4. System Overview - Integration of Image Segmentation Maps Using Region and Edge Information.

This algorithm is not another segmentation algorithm, but is an integration algorithm capable of using all established segmentation algorithms as *front ends*. Such an integration algorithm allows many fast (but not necessarily intelligent) segmentation algorithms to be used in parallel to achieve fast segmentation, and their results to be fed into the integration module. For example, multiple *shape-from-X* techniques can operate in parallel and then have their results integrated. Consequently, the need to design a single superior algorithm for segmentation is much less critical. Since only local information is used in the initial estimation and in the contour smoothing stages, the algorithm may be ported to parallel/distributed computing hardware when maximum operating speed is required.

b. Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot.

We have developed a new approach to extracting important line segments from monocular images in order to estimate the position of important objects in the path of an autonomous robot [21-24]. A paper based upon this work received the Phillips Award for Best Paper at the IEEE Computer Society International Conference on Robotics and Automation, Nice, France (1992). All stages of image interpretation, including the lowest processing level, are designed to provide the higher stages with the most semantically useful features.

The robot's tasks are usually specified in high-level semantic terms, such as "go down the hallway and go through the last door on the left." In order to execute this task, the robot must be able to identify the objects of interest—hallways and doors—in its perception of the environment. One approach is to reconstruct 3D line segment descriptions of the environment from several intensity images, which are then grouped and matched to selected object models. For this, several 3D segments are selected according to orientation and position criteria. The basic idea is that the architecture in most indoor scenes contains edges with particular orientations in 3D, such as vertical and two horizontal orientations perpendicular to each other. In a 2D perspective projection, all the edges with a given 3D orientation appear to converge to a single point, called the vanishing point (Figure 5). By precomputing the position of vanishing points in each image, it is

possible to find the most likely 3D orientation of observed edges, even from a single monocular image. A 3D hypothesis of an object is generated and matched to the selected segments. The process is repeated for other segments and other hypotheses. The high level semantic interpretation is then used to determine the free space or to find objects of interest.

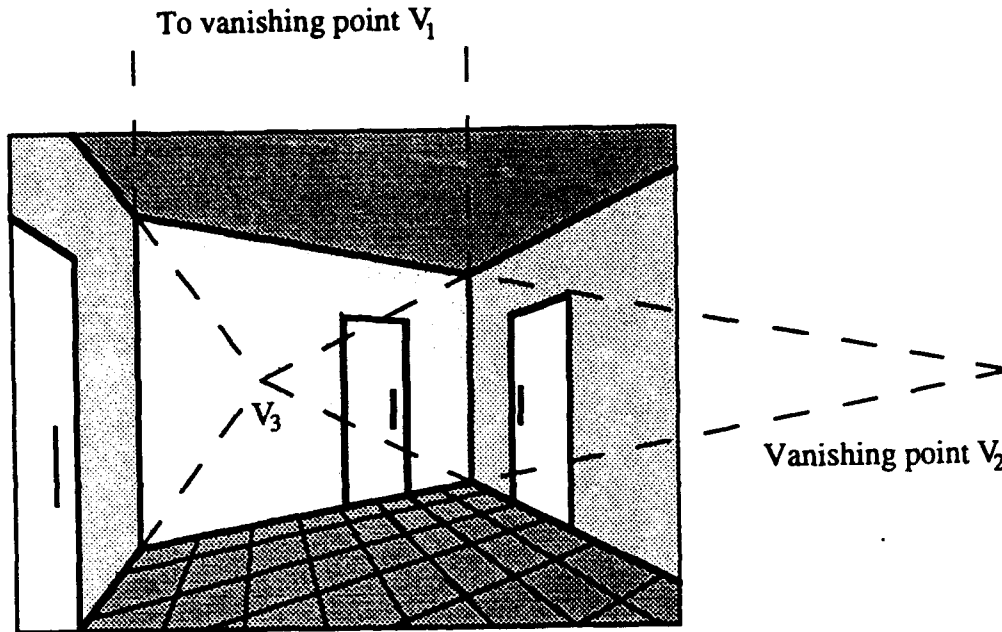


Figure 5. Perspective Projection of a Scene

Since high-level interpreters are looking for line segments of particular orientations in 3D, we designed the lower level image-processing stages to take advantage of this information. This top-down information can benefit the feature extraction stage by reducing the number of unwanted features, increasing sensitivity to good features, and drastically speeding the computation. Preliminary results of this work were presented at the 1992 IEEE Intl. Conference on Robotics and Automation, in a paper that received the Best Student Paper Award of the conference [21].

c. 3D Structure Reconstruction from an Ego Motion Sequence Using Statistical Estimation and Detection Theory.

We have derived a decision-theoretical algorithm to estimate 3D structures from extended sequences of 2D images taken by a moving camera [25]. We assume that the camera motion is known and that the world is stationary. The 3D structures of interest are 3D lines, because they are relatively stable and easy to extract from images. Traditionally, feature-based motion analysis involves several separate operations: feature detection, feature matching, structure/motion estimation, and higher level processing, such as feature grouping. Most of these operations were originally designed to operate on just one or two images, and thus they did not take advantage of having an extended sequence of images. The decision-theoretical algorithm, however, uses statistical estimation and detection theory to integrate these operations. Statistical estimation theory has been used extensively in sequential 3D structure reconstruction problems. The basic idea is that if the 3D world remains stationary, then the unknown 3D structures, such as depths or line parameters, can be considered as state variables; the changes in their values due to the camera's ego-motion can be modeled by the system's dynamics, and the information ex-

tracted from images is related to the unknowns via the so-called observation or measurement models. The structure estimation can thus be posed as the state reconstruction problem in dynamic system theory. The novelty of our work is that the statistical detection theory is used to model the other phases of the operation, and thus provides a natural way to incorporate temporal information into those phases and to integrate the whole system. This formulation is more robust against matching errors because it represents the matching ambiguities by their probabilities and explicitly incorporates the information into the estimation update. Feature tracking and parameter estimation are closely coupled in this formulation. The recursive estimation algorithm provides the prediction used for tracking features, and the tracking algorithm evaluates the probabilities used for updated estimation. This algorithm could be easily modified to include other properties associated with lines, such as length, midpoint, or contextual information associated with 2D edges, as long as they can be represented in a parametric form. The algorithm could also be modified for other types of features, such as points or curves, as long as they can be represented by a finite number of parameters.

d. Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images.

Our group has developed a new algorithm for matching line segments based on perceptual grouping relaxation labeling [26]. We consider feature matching between two views as a "temporal grouping" process, in addition to the traditional spatial groups established by perceptual grouping in a single images. The relaxation labeling paradigm is used to integrate spatial and temporal grouping. In the relaxation process, correspondence ambiguities are resolved by iteratively propagating constraint information from the nodes of a line, which are defined by the line's perceptual groups. The system is suitable for indoor or urban structured environments. This line matching algorithm has three new developments: (1) rigorous perceptual grouping processes based on the statistical inference paradigm; (2) generalization of the perceptual grouping concept to the temporal domain; and (3) the cooperation of spatial and temporal grouping processes using relaxation labeling techniques. We detect edges using the Canny edge operator and then extract lines using the Object Recognition Toolkit. The algorithm uses the relaxation labeling paradigm to match a set of lines in two images. Initially, a line in the first image has multiple matching candidates in the second image. This ambiguity is resolved by iteratively propagating constraint information from a line's neighbors, which are defined in the iterative update process by the line's perceptual groups. Traditional perceptual grouping requires heuristic threshold values. We define the grouping process using the statistical inference paradigm. The advantage of this algorithm over previous work is that we simultaneously hypothesize and test both temporal and spatial relations among 2D lines, allowing one relation to be used as supporting evidence for the other.

e. Stereo Image Interpretation in the Presence of Narrow Occluding Objects.

In a study of the use of stereo vision to establish object correspondence, we reviewed major developments in establishing stereo correspondence for the extraction of 3-D structure of a scene; identified broad categories of algorithms based upon differences in image geometry, matching primitives, and computational structure; and reviewed the performance of these stereo techniques on various test images [27].

In this study, we examined the twin issues of the gain in accuracy of stereo correspondence and the accompanying increase in computational cost due to the use of a third camera for stereo analysis [28]. Trinocular stereo algorithms differ from binocular algorithms in the epipolar constraint used in the local matching stage. The current literature does not provide any insight into the relative merits of binocular and trinocular stereo matching with the matching accuracy being verified against the ground truth. We conducted experiments to evaluate the relative

performance of binocular and trinocular stereo algorithms using stereo images generated by applying a Lambertian reflectance model to real digital elevation maps (DEMs) from the U. S. Geological Survey. The matching accuracy of the stereo algorithms was evaluated by comparing the observed stereo disparity against the ground truth derived from the DEMs. We observed that trinocular local matching reduced the percentage of mismatches having large disparity errors by more than half, as compared to binocular matching, but increased the computational cost by only about one fourth. We also performed a quantization-error analysis of the depth reconstruction process for the nonparallel stereo-imaging geometry used in the experiments

We have developed an approach to stereo vision that utilizes the Dynamic Disparity Search (DDS) framework, which combines the spatial hierarchy with a new disparity hierarchy mechanism to reduce stereo matching errors caused by the presence of narrow occluding objects [29-30]. Narrow, occluding objects in stereo images cause matching errors that cannot be handled by the spatial hierarchy method alone. The merits of the DDS approach are demonstrated on real stereo images.

Most contemporary stereo correspondence algorithms impose global consistency among candidate match-points using only a spatial hierarchy mechanism. As mentioned above, the spatial hierarchy mechanism cannot handle the presence of narrow occluding objects. We have analyzed the stereo matching failures caused by the spatial hierarchy mechanism [31] and formulated a new global matching framework. Experimental results show a significant decrease in the false positive match-rate using this framework.

f. Analysis of Video Images Using Point and Line Correspondences.

We have investigated the problem of analyzing time-varying imagery using a feature-based approach [32]. We assume a scenario in which the imaged objects remain stationary while the camera moves. The goal is to compute the structure of the imaged objects and the motion of the camera from a sequence of video images. Our method exploits the principle of the invariance of rigid configurations during motion. Using the rigidity constraints, we specify equations based on distance and angular invariance to compute the structural parameters of the imaged 3-D objects independently of the camera's motion. Once the structural parameters are recovered, the motion parameters can be computed. The advantage of this approach is in the decomposition of the computations of structure and motion, and the simultaneous use of point and line correspondences, which allows our approach to not be limited to objects whose images have only a particular type of feature in abundance. Our study considers the special case of four points and one line, which is the minimum feature set required for computing the structure and motion parameters, as well as additional features in an overdetermined system of equations to improve the reliability and accuracy of the computation. The algorithm's validity was demonstrated using computer simulation results as well as results from real image sequences.

## 4. Autonomous Navigation.

### a. Construction of an Autonomous Mobile Robot, Robo-Tex.

Mobile robots are finding an increasing number of applications in military and civilian environments. Tele-operated robots, guided by remotely located human operators, now perform many operations in hazardous environments such as high-radiation zones. Many more potential applications exist in manufacturing, surveillance and planetary exploration. However, to be truly useful, robots will need to be more autonomous in perceiving, understanding and responding to the environment. Computer vision--the automated understanding of video images and other sensor data--is central to the development of robots that can self-navigate and perform useful tasks in indoor and outdoor environments.

Although it is fairly easy to link video cameras and computers, the automatic understanding of digitized images is still a very difficult task. Traditionally, researchers have used several cameras in a stereo-vision setup to perceive the depth of objects. Image features are extracted from each image and matched with one another, then depth is computed by triangulation. The resulting 3-dimensional representation of the robot's environment is used by path-planning algorithms to navigate the robot and avoid obstacles in its path.

An innovative mobile robot, Robo-Tex, has been developed at the Computer and Vision Research Center [33-34]. As pictured in Figure 6, Robo-Tex uses a single video camera and perceives depth by tracking features over a sequence of images. This simplifies the robot hardware and reduces the amount of computation needed for image processing.

One common problem in processing image data is the selection of significant, meaningful features from the image and the elimination of less meaningful features. In an indoor, man-made environment such as an office building, some prominent features that are useful for navigation are the boundaries, or edges, of the walls and doorways that the robot must identify in order to navigate successfully. In man-made environments, most of the significant edges (those corresponding to walls and doorways) have particular 3-dimensional orientations, usually vertical and horizontal. As described in Section 3 above, we have developed new perception algorithms for Robo-Tex that concentrate on such edges, thereby eliminating many insignificant features [21-24].

Robo-Tex relies on the geometrical properties of vanishing points to estimate the most likely orientation of the edges in each 2-dimensional image. For example, in a 3-dimensional scene, vertical edges appear to converge to one point, the vanishing point of vertical lines. This approach reduces the number of unwanted features, increases the sensitivity to useful features, and drastically speeds the computation.

While most mobile robots are connected to large computers by cable or radio links, Robo-Tex carries an onboard HP-735 workstation. The workstation is easier to program than the special-purpose boards usually used for robot control, yet it has the computing power required by vision algorithms. This configuration allows new vision algorithms to be tested quickly and easily. Electrical power for the workstation and other onboard equipment is provided by 12 V batteries and a 110 V AC power inverter. Standard equipment can be added to the robot by simply plugging it in.

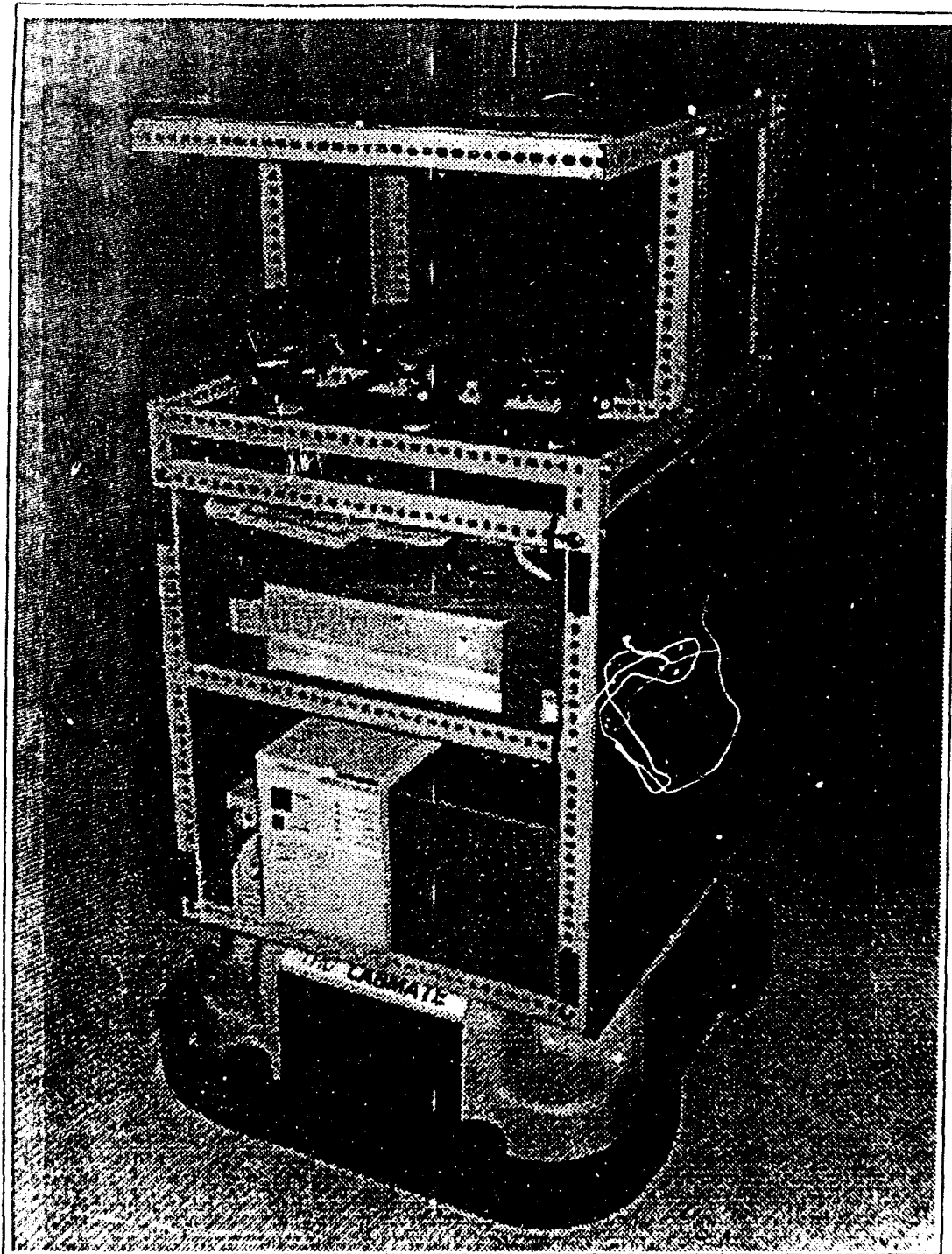


Figure 6. The mobile robot, Robo-Tex.

The ultimate goal of Robo-TeX is to navigate autonomously in both indoor and outdoor environments while building an accurate CAD model of the world. Already, significant progress toward that goal has been made. The robot can visually measure distances between edges with a precision comparable to that available from an architect's plan and automatically generate complete CAD models of buildings [32-33]. The techniques used in this robot vision system find application in a number of areas, including architecture and graphics, robot navigation, active vision, and scene understanding.

The Robo-TeX vision system uses a 3D representation of the environment that concentrates on architecturally significant features, and which is more accurate than is strictly necessary for navigation. In order to create an architectural CAD model of the environment, the vision system must not represent a large number of insignificant details. In designing this system, the objective was to design all stages of image interpretation, including the lowest image processing levels, to provide higher stages with the most semantically useful features. The system includes a line segment detector (described above [18-21]), an automatic tracker, and a CAD modeler optimized for environments with prominent 3D orientations [30-31].

b. Position Estimation Techniques for an Autonomous Mobile Robot.

The development of truly autonomous mobile robots is one of the most challenging and important applications of computer vision. The basic tasks involved in robot navigation are as follows: (1) sensing the environment; (2) mapping the environment, e.g., building a representation of the environment; (3) locating itself with respect to the environment (position estimation); and (4) planning and executing efficient routes in the environment (path planning and obstacle avoidance). Several projects have been carried out under this contract to develop more efficient and accurate methods of position estimation. Position estimation techniques vary, depending upon the environment in which the robot must navigate (indoor or outdoor), the type of robot sensors used (visual, range, etc.), and the information that is known about the environment (map representation, coordinate systems, etc.). Position estimation methods can be broadly classified into four categories: landmark-based methods; methods using trajectory integration and dead reckoning; methods using a standard reference pattern; and methods using *a priori* knowledge of a world model matched to sensor data for position estimation. The two projects described below fall into the last category.

In methods that match sensor data to a world model, the model (or map) of the environment may be a CAD description of the environment, a floor map, or, in outdoor terrain, a digital elevation map (DEM). To estimate the robot's position in the environment, the robot's sensor observations are matched to the given map. Once a correspondence is established between the sensor data and the world model (map) data, the robot's position and pose is calculated as a co-ordinate transformation that transforms the world model into the sensor co-ordinate system.

Although mobile robots are equipped with wheel encoders that can estimate the robot's position at every instant, these estimates contain errors due to wheel slippage and quantization effects. As the robot moves, these errors accrue and can grow limitlessly, causing the position estimate to become increasingly uncertain. Therefore, most mobile robots use additional forms of sensing, such as vision, to aid the position estimation process.

1) Position Estimation of an Autonomous Mobile Robot In Cluttered Outdoor Environments Using Geometric Visibility Constraints. In this project, we have considered the position estimation of an autonomous mobile robot navigating in an outdoor urban, man-made environment consisting of polyhedral buildings with flat rooftops [37-42]. Our world model is made up of the 3-D descriptions of the line segments that compose the buildings' rooftops. Our robot sensor is a video camera mounted on the robot. Establishing correspondence between the image and the map is particularly difficult in this case, since they are of different dimensionality

(a 2-D image and a 3-D map), in different formats, and are described in different coordinate frames. Nevertheless, optical imaging is still the preferred form of sensing since it is totally passive.

Our procedure is to extract features from the image and then search the map to locate the corresponding features. For a large map, an exhaustive search would require an enormous amount of computation. To reduce the computational cost, we limit the search using visibility constraints imposed by the model's geometry and the known camera geometry. We use a two-stage constrained-search strategy. Stage One is a coarse search that narrows the robot location to a small set of possible locations; Stage Two searches exhaustively through this set and accurately establishes the robot's position in the environment.

The use of visibility constraints is a novel method that captures the geometric constraints between the 3-D model and the 2-D image features by using a new, viewer-centered, intermediate representation of the robot's environment from the given world model, called edge visibility regions (EVRs). The EVR inherently captures the geometric visibility constraints between the world model and image features by partitioning the ground plane (the plane in which the robot navigates) into a set of distinct, non-overlapping regions. Each region has an associated list of the world model features visible in that region, known as the Visibility List (VL). Thus, each EVR is a region of space with the topological property that from its points the same set of edges of the model are visible through a complete circular scan. These geometric constraints are compiled off-line, thereby reducing the runtime of the position estimation process.

Once the EVR description of the environment is formed off-line using the given world model description, we use a modified Hough transform technique to perform transform clustering and isolate the set of EVRs that are most likely to contain the robot location. In addition, we can efficiently reduce the complexity of the search by propagating the geometric constraints established by the EVRs to correctly identify the robot's position. These search techniques have proven to be very robust to image feature detection errors such as missing features and spurious features. The method is also robust to incomplete model descriptions and inaccurate feature representation. Further, the EVR representation of the environment is advantageous in the mobile robot's path-planning tasks.

2) Position Estimation Of An Autonomous Mobile Robot In An Outdoor Natural, Mountainous Environment. While the problems associated with navigating mobile robots in an indoor structured environment are reasonably well studied and a number of different approaches have been suggested, outdoor navigation of a mobile robot in an unstructured environment is a more complex problem and many issues remain open and unsolved.

To consider the problem of estimating the position of an autonomous mobile robot navigating in an outdoor mountainous environment, we assume that the robot is provided with a visual camera that can be panned and tilted, as well as a compass and an altimeter to measure its altitude [43-44]. A Digital Elevation Map (DEM) of the navigation area is provided. The DEM is a 3-D database that records the terrain elevations for ground positions at regularly spaced intervals. Our problem is to find common features to match the 2-D images intensity images from the robot camera to the 3-D DEM, and thereby estimate the robot's position.

Our approach to this problem is to extract features from the images and then search the map for corresponding features. Once this correspondence is found, the position can then be computed. As with the project described earlier, an exhaustive search of the entire map would be prohibitively expensive. So, we formulate this correspondence problem as a constrained search problem. Since the robot is assumed to be located in the DEM, the DEM grid is used as a quantized version of the entire space of possible robot locations. The feature used to search the DEM is the shape and position of the horizon line contour (HLC) in the image plane. From the current

robot position, images are taken in the four geographic directions, N,S,E, and W. The contour of the horizon line (HLC) is extracted from these images and coded. Using the height of the contour line in the image plane and the known camera geometry as input parameters, the entire DEM is searched for possible camera locations so that the points in the elevation map project onto the image plane to form a contour of the shape and height we are searching for. Since searching the DEM exhaustively for the exact shape of the horizon line is a very computationally intensive process, we split the search into two stages. In stage 1, we search using the height of the horizon line at the center of the image plane in all the four images. Geometric constraints derived from the camera geometry and the height of the HLC are used to prune large subspaces of the search space; finally, the position is isolated to a small set of possible locations. In stage 2, these locations are then considered as the candidate robot positions, and the actual image that would be seen at these points is generated using computer graphics rendering techniques from the DEM. The HLCs are also extracted from these images and then compared with the original HLCs to arrive at a measure of the new HLC's disparity. The robot location corresponding to the lowest disparity is then considered as the best estimate of the robot's position. As the results show, the approach is quite effective and, in almost all cases, the position estimate is very close to the actual position.

c. Calibrating a Mobile Camera's Parameters.

Our research has addressed the problem of the calibration of the relative rotation and translation between a camera and a mobile robot's coordinate system, as well as the camera's intrinsic parameters, from a sequence of monocular images and robot movements [45-46]. Existing hand/eye calibration procedures for robot arms are not directly applicable because they require the robot hand to have at least two rotational degrees of freedom. A suitable representation for camera rotation is used, and the calibration task is decomposed into two stages. Furthermore, to recover the camera's rotation motion, the inverse perspective geometry constraints of a rectangular corner are employed. Complicated calibration patterns are thus not needed. The calibration procedures were tested using both synthetic and real data.

d. Calibration Procedure for a Fish-Eye Lens (Super-Wide-Angle Lens) Camera.

Calibration of cameras is an important issue in computer vision. Accurate camera calibration is crucial in applications that involved quantitative measurements, such as 3-D sensing and measurement for robotic vision. Because they provide an extremely large field of view, fish-eye lenses are useful in robotic vision for allowing close objects to be viewed in their entirety and for perceiving objects that appear from an unpredictable direction. We have developed a new algorithm for the geometric camera calibration of a fish-eye lens mounted on a CCD TV camera [47]. The algorithm determines a mapping between points in the world coordinate system and their corresponding point locations in the image plane. The parameters to be calibrated are the effective focal length, one-pixel width on the image plane, image distortion center, and distortion coefficients. A simple calibration pattern consisting of equally spaced dots is introduced as a reference for calibration. Some parameters to be calibrated are eliminated by setting up the calibration pattern precisely and assuming negligible distortion at the image distortion center. Thus, the number of unknown parameters to be calibrated is drastically reduced, enabling simple and useful calibration. The method employs a polynomial transformation between points in the world coordinate system and their corresponding image plane locations. The coefficients of the polynomial are determined using the Lagrangian estimation. The effectiveness of the proposed calibration method is confirmed by experimentation.

## CONCLUSIONS.

The work outlined above has demonstrated amply that the fusion of multiple sensing modalities has much to contribute to machine vision. We have continued to make significant advances in the development of algorithms and techniques for fusing laser radar and thermal images, and have addressed one of the most difficult areas of computer vision, namely, object recognition in cluttered scenes. Significant strides have also been made in the development of techniques for autonomous navigation. This work has been enthusiastically received by the computer vision research community. Our results have been presented at reviewed conferences such as the IEEE International Conference on Robotics and Automation, the IEEE International Workshop on Intelligent Robotics and Systems, and the IEEE Conference on Computer Vision and Pattern Recognition, and published in refereed journals including *Pattern Recognition*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Machine Vision and Applications*, *Computer Vision, Graphics and Image Processing: Image Understanding*, and *IEEE Transactions on Robotics and Automation*. Book chapters based on this research have appeared in the following books: *Multisensor Fusion for Computer Vision*; *The Handbook of Pattern Recognition and Computer Vision*; *Encyclopedia of Artificial Intelligence*; *Control and Dynamic Systems: Advances in Theory and Applications, Volume 39: Advances in Robotic Systems*; *Parallel Processing for Artificial Intelligence*, and *Autonomous Mobile Robots: Perception, Mapping, and Navigation*.

Three papers based on research under this contract have received distinguished awards, including "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," (IEEE Computer Society Outstanding Paper Award, 7th Conference on Artificial Intelligence Applications (1991) [1]), "Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot," (Phillips Award for Best Paper at the IEEE Computer Society International Conference on Robotics and Automation, Nice, France (1992) [21]), and "Applying Perceptual Organization to the Detection of Man-made Objects in Non-Urban Scenes," (Honorable Mention of the Pattern Recognition Society Award for Outstanding Contribution, November 1993, [8]). A total of 13 graduate and 4 undergraduate students were supported under this contract, and 6 Ph.D. and 1 M.S. degrees were completed during the contract term.

Although we have made significant progress in the fusion of multiple sensing modalities for machine vision, the development of truly general-purpose machine vision systems that are capable of true autonomy in sensing, understanding, and responding to their environment remains a distant objective. To further progress toward that goal, we must examine the techniques employed by human vision, which is significantly better than machine vision at detecting and recognizing objects because human vision integrates information from a number of sources and apparently uses a number of different "algorithms." Current approaches to computer/machine vision use a number of different techniques, including statistical pattern recognition, model-based vision, neural networks, knowledge-based artificial intelligence (rule-based systems), and adaptive and learning systems. Despite much investigation, machine vision techniques, including image segmentation and multisensor fusion, have had only limited success in identifying three-dimensional objects in cluttered environments. One possible means of improving performance is to integrate multiple vision techniques into one system. For example, model based vision techniques are effective in situations where the models of both the objects to be recognized and the environment are well understood. On the other hand, an approach based on artificial neural networks (ANNs) is effective when model-based knowledge is unavailable or the number of variables is so large as to make a model-based strategy unworkable. An intelligent hybrid vision system that combines these two techniques could be successful in situations in which systems based only model-based vision or only ANNs would fail. Further, not only can humans integrate information despite irrelevant and incomplete information, they can integrate spatial and

temporal information, as well as integrate contextual information with model-based knowledge. This suggests that the integration of temporal and spatial information could further assist such a hybrid vision system. The superiority of human vision appears to lie in its use of multiple sources of information and processing techniques, leading us to believe that a vision system using a similar strategy to combine model-based vision with ANNs on spatial/temporal images should be investigated.

Vision systems capable of recognizing 3D objects in a natural environment often encounter significant problems, including noise, occlusion, low contrast, low resolution, and background clutter. Given the knowledge we have today, this problem may not be immediately solvable. However, through further exploration of combining existing techniques, substantial progress toward that goal can be made.

## BIBLIOGRAPHIC REFERENCES.

- [1] Chu, Chen-Chau, and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," *Proc. 7th IEEE Conference on Artificial Intelligence Applications*, pp. 190-196, February 1991.
- [2] Chu, Chen-Chau and J. K. Aggarwal, "The Interpretation of Laser Radar Images by a Knowledge-Based System," *Machine Vision and Applications*, Vol. 4, pp. 145-163, 1991.
- [3] Chu, C. C. and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," Computer and Vision Research Center Technical Report RT-91-6-70, March 1991.
- [4] Chu, C. C. and J. K. Aggarwal, "Image Interpretation Using Multiple Sensing Modalities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 840-846, 1992.
- [5] Chu, Chen-Chau and J. K. Aggarwal, "The Integration of Image Segmentation Maps Using Region and Edge Information," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 12, pp. 1241-1252, December 1993.
- [6] Chu, C. C., J. Ghosh, and J. K. Aggarwal, "On Supporting Rule-Based Image Interpretation Using a Distributed Memory Multicomputer," *Parallel Processing for Artificial Intelligence*, L. Kanal, et al., Editors, Elsevier Science Publishers, B.V., in press.
- [7] Lu, H. Q. and J. K. Aggarwal, "Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes," Computer and Vision Research Center Technical Report RT-91-5-69, March 1991.
- [8] Lu, H. Q., and J. K. Aggarwal, "Applying Perceptual Organization to Detection of Man-made Objects in Non-Urban Scenes," *Pattern Recognition*, pp. 835-853, 1992.
- [9] Lu, H. Q., and J. K. Aggarwal, "Extracting Geometric Structures and a Region of Interest for Man-made Object Detection," *Proc. 20th AIPR Workshop: Computer Vision--Meeting the Challenges*, McLean, Virginia, October 1991.
- [10] Nandhakumar, S. Karthik, and J. K. Aggarwal, "Unified Modeling of Non-Homogeneous 3-D Objects for Thermal and Visual Image Synthesis," *Pattern Recognition*, in press.
- [11] Nair, Dinesh, and J. K. Aggarwal, "An Object Recognition System Using ART-2," *Proc. 3rd International Conference on Advances in Pattern Recognition and Digital Techniques*, December 28-31, 1993, Calcutta, India
- [12] Sabata, Bikash and J. K. Aggarwal, "Correspondence of Surfaces in a Sequence of Range Images for Motion Estimation and Tracking," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 385-388.
- [13] Sabata, Bikash and J. K. Aggarwal, "Hypergraph Based Feature Matching in a Sequence of Range Images," *Proceedings of the 4th International Workshop on Time Varying Image Processing and Moving Object Recognition*, Florence, Italy, June 10-11, 1993.

- [14] Sabata, Bikash and J. K. Aggarwal, "Surface Correspondence and Motion Computation from a Sequence of Range Images" *1994 IEEE International Conference on Robotics and Automation*, San Diego, CA, May 8-13, 1994, pp. 1052-1059.
- [15] Sabata, B., F. Arman, and J. K. Aggarwal, "Segmentation of 3D Range Images Using Pyramidal Data Structures," *Computer Vision, Graphics and Image Processing: Image Understanding*, Vol. 57, No. 3, 1993, pp. 373-387.
- [16] Sabata, B., F. Arman, and J. K. Aggarwal, "The Convergence of Fuzzy Pyramid Algorithms," *J. Mathematical Imaging and Vision*, in press.
- [17] Arman, F., and J. K. Aggarwal, "CAD-Based Object Recognition in 3-D Range Images Using Pre-compiled Recognition Strategy," Computer and Vision Research Center Technical Report TR-92-11-86.
- [18] Arman, Farshid, and J. K. Aggarwal, "Generating Automatic Recognition Strategies," *Proc. IEEE Workshop on Directions in Automated CAD-Based Vision*, Hawaii, pp. 124-133, June 1991.
- [19] Arman, F., and J. K. Aggarwal, "Model-Based Object Recognition in Dense-Range Images--A Review," *ACM Computing Surveys*, Vol. 25, No. 1, March 1993.
- [20] Arman, F., and J. K. Aggarwal, "CAD-Based Vision: Object Recognition in Cluttered Range Images Using Recognition Strategies," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 58, no. 1, pp. 33-48, July 1993.
- [21] Lebègue, Xavier and J. K. Aggarwal, "Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot," *Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, May 1992, pp. 1778-1785.
- [22] Lebègue, Xavier and J. K. Aggarwal, "Semantically Significant Line Segments for a Mobile Robot--ROBOTEX," *Video Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, May 1992. (Videotape presentation).
- [23] Lebègue, Xavier and J. K. Aggarwal, "Detecting 3-D Parallel Lines for Perceptual Organization," *Proceedings of the Second European Conference on Computer Vision*, May 1992, pp. 720-724.
- [24] Lebègue, Xavier, and J. K. Aggarwal, "Significant Line Segments for an Indoor Mobile Robot," *IEEE Trans. on Robotics and Automation*, Vol. 9, No. 6, December 1993, pp. 801-815.
- [25] Chang, Yuh-Lin, and J. K. Aggarwal, "3D Structure Reconstruction from an Ego Motion Sequence Using Statistical Estimation and Detection Theory," *Proc. IEEE Workshop for Visual Motion*, Princeton, New Jersey, October 1991, pp. 268-273.
- [26] Chang, Yuh-lin, and J. K. Aggarwal, "Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images," *Proc. of the Intl. Conf. on Digital Signal Processing and 2nd Intl. Conf. on Computer Applications to Engineering Systems*, pp. 622-627, July 14-16, 1993, Nicosia, Cyprus.
- [27] Dhond, Umesh R. and J. K. Aggarwal, "Structure from Stereo--A Review," reprinted in *Autonomous Mobile Robots: Perception, Mapping, and Navigation*, S. Sitharama

Iyengar and Alberto Elfes, Editors, IEEE Computer Society Press, Los Alamitos, California, 1991, pp. 25-46.

- [28] Dhond, U. R., and J. K. Aggarwal, "A Cost-Benefit Analysis for a Third Camera for Stereo Correspondence," *International Journal of Computer Vision*, Vol. 6, No. 1, pp. 39-58, 1991.
- [29] Dhond, U., and J. K. Aggarwal, "Stereo Image Interpretation in the Presence of Narrow Occluding Objects," Computer and Vision Research Center Technical Report TR-92-13-88.
- [30] Dhond, Umesh and J. K. Aggarwal, "Computing Stereo Correspondences in the Presence of Narrow Occluding Objects," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, 1992, pp. 758-764.
- [31] Dhond, Umesh and J. K. Aggarwal, "Analysis of the Stereo Correspondence Process in Scenes with Narrow Occluding Objects," *Proceedings of the 11th International Conference on Pattern Recognition*, The Hague, The Netherlands, August 30 - September 3, 1992, Vol. I, pp. 470-473.
- [32] Wang, Y. F., N. Karandikar, and J. K. Aggarwal, "Analysis of Video Image Sequences Using Point and Line Components," *Pattern Recognition*, Vol. 24, No. 11, pp. 1065-1084, 1991.
- [33] Lebègue, X. and J. K. Aggarwal, "A Mobile Robot for Visual Measurements in Architectural Applications," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 195-198.
- [34] Lebègue, X. and J. K. Aggarwal, "ROBOTEX: An Autonomous Mobile Robot for Precise Surveying," *Proc. International Conference on Intelligent Autonomous Systems (IAS-3)*, February 15-19, 1993, pp. 460-469.
- [35] Lebègue, Xavier and J. K. Aggarwal, "Generation of Architectural CAD Models Using a Mobile Robot," *Proc. 1994 IEEE Intl. Conf. on Robotics and Automation*, San Diego, CA, pp. 711-717.
- [36] Lebègue, Xavier, and J. K. Aggarwal, "Automatic Creation of Architectural CAD Models," *Proc. 2nd IEEE Computer Society CAD-Based Vision Workshop*, Seven Springs, Pennsylvania, Feb. 8-10, 1994.
- [37] Talluri, R. J., and J. K. Aggarwal, "Edge Visibility Regions - A New Representation of the Environment of a Mobile Robot," *Proc. IAPR Workshop on Machine Vision Applications*, Tokyo, 1990, pp. 375-380. (reprints mailed 2/25/91.)
- [38] Talluri, Raj, and J. K. Aggarwal, "Position Estimation of a Mobile Robot using Edge Visibility Regions," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, pp. 714-715, June 1991.
- [39] Talluri, R. J. and J. K. Aggarwal, "Mobile Robot Self-Location Using Constrained Search," *Proc. IEEE International Workshop on Intelligent Robotics and Systems*, Osaka, Japan, November 1991.

- [40] Talluri, R. and J. K. Aggarwal, "Transform Clustering for Model-Image Feature Correspondence," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 579-582.
- [41] Talluri, R. and J. K. Aggarwal, "Autonomous Navigation in Cluttered Outdoor Environments using Geometric Visibility Constraints," *International Conference on Intelligent Autonomous Systems (IAS-3)*, February 15-19, 1993, pp. 186-195.
- [42] Talluri, R. and J. K. Aggarwal, "Position Estimation Techniques for an Autonomous Mobile Robot - A Review," in *The Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Editors, World Scientific Publishing Co., Inc., Singapore, 1993, pp. 769-801.
- [43] Talluri, R. and J. K. Aggarwal, "Position Estimation for an Autonomous Mobile Robot in an Outdoor Environment," *IEEE Trans. on Robotics and Automation*, Vol. 8, No. 5, pp. 573-584, 1992
- [44] Talluri, R. and J. K. Aggarwal, "Image/Map Correspondence for Mobile Robot Self-location Using Computer Graphics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Special Issue on 3D Modeling in Image Analysis And Synthesis, Vol. 15, No. 6, June 1993, pp. 597-601.
- [45] Chang, Yuh-Lin, and J. K. Aggarwal, "Calibrating a Mobile Camera's Extrinsic Parameters with Respect to Its Platform," *Proc. IEEE International Symposium on Intelligent Control*, Arlington, VA, August 1991.
- [46] Chang, Yuh-Lin, Xavier Lebègue, and J. K. Aggarwal, "Calibrating a Mobile Camera's Parameters," *Pattern Recognition*, Vol. 26, No. 1, pp. 75-88, 1993.
- [47] Shah, Shishir and J. K. Aggarwal, "A Simple Calibration Procedure for Fish-Eye (High Distortion) Lens Camera," *Proc. 1994 IEEE Intl. Conf. on Robotics and Automation*, San Diego, CA, pp. 3422-3427.

## PUBLICATIONS AND TECHNICAL REPORTS.

Aggarwal, J. K., and C.C. Chu, "The Issues, Analysis and Interpretation of Multisensor Images," in *Multisensor Fusion for Computer Vision*, J. K. Aggarwal, Editor, Springer Verlag, Berlin, 1993, pp. 37-62.

Aggarwal, J. K., B. Sabata, and F. Arman, "Range Images: Segmentation, Recognition Of Objects, And Motion Estimation," Invited Talk, 3rd International Conference on Advances in Pattern Recognition and Digital Techniques, December 28-31, 1993, Calcutta, India (P. C. Mahalanobis Birth Centenary Celebration).

Aggarwal, J. K., and Y. F. Wang, "Sensors and Sensor Fusion," *Encyclopedia of Artificial Intelligence*, Volume 2, Stuart C. Shapiro, Editor-in-Chief, John Wiley & Sons, Inc., 1992, pp. 1511-1526.

Arman, F., and J. K. Aggarwal, "CAD-Based Object Recognition in 3-D Range Images Using Pre-compiled Recognition Strategy," Computer and Vision Research Center Technical Report TR-92-11-86.

Arman, F., and J. K. Aggarwal, "CAD-Based Vision: Object Recognition in Cluttered Range Images Using Recognition Strategies," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 58, no. 1, pp. 33-48, July 1993.

Arman, Farshid, and J. K. Aggarwal, "Generating Automatic Recognition Strategies," *Proc. IEEE Workshop on Directions in Automated CAD-Based Vision*, Hawaii, pp. 124-133, June 1991.

Arman, F., and J. K. Aggarwal, "Model-Based Object Recognition in Dense-Range Images--A Review," *ACM Computing Surveys*, Vol. 25, No. 1, March 1993.

Chang, Yuh-Lin, and J. K. Aggarwal, "3D Structure Reconstruction from an Ego Motion Sequence Using Statistical Estimation and Detection Theory," *Proc. IEEE Workshop for Visual Motion*, Princeton, New Jersey, October 1991, pp. 268-273.

Chang, Yuh-Lin, and J. K. Aggarwal, "Calibrating a Mobile Camera's Extrinsic Parameters with Respect to Its Platform," *Proc. IEEE International Symposium on Intelligent Control*, Arlington, VA, August 1991.

Chang, Yuh-Lin, Xavier Lebègue, and J. K. Aggarwal, "Calibrating a Mobile Camera's Parameters," *Pattern Recognition*, Vol. 26, No. 1, pp. 75-88, 1993.

Chang, Yuh-lin, and J. K. Aggarwal, "Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images," *Proc. of the Intl. Conf. on Digital Signal Processing and 2nd Intl. Conf. on Computer Applications to Engineering Systems*, pp. 622-627, July 14-16, 1993, Nicosia, Cyprus.

Chu, C. C. and J. K. Aggarwal, "Image Interpretation Using Multiple Sensing Modalities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 840-846, 1992.

Chu, C. C. and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," Computer and Vision Research Center Technical Report RT-91-6-70, March 1991.

- Chu, Chen-Chau and J. K. Aggarwal, "The Interpretation of Laser Radar Images by a Knowledge-Based System," *Machine Vision and Applications*, Vol. 4, pp. 145-163, 1991.
- Chu, Chen-Chau, and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," *Proc. 7th IEEE Conference on Artificial Intelligence Applications*, pp. 190-196, February 1991.
- Chu, Chen-Chau and J. K. Aggarwal, "The Integration of Image Segmentation Maps Using Region and Edge Information," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 12, pp. 1241-1252, December 1993.
- Chu, C. C., J. Ghosh, and J. K. Aggarwal, "On Supporting Rule-Based Image Interpretation Using a Distributed Memory Multicomputer," *Parallel Processing for Artificial Intelligence*, L. Kanal, et al., Editors, Elsevier Science Publishers, B.V., in press.
- Dhond, U. R., and J. K. Aggarwal, "A Cost-Benefit Analysis for a Third Camera for Stereo Correspondence," *International Journal of Computer Vision*, Vol. 6, No. 1, pp. 39-58, 1991.
- Dhond, U., and J. K. Aggarwal, "Stereo Image Interpretation in the Presence of Narrow Occluding Objects," Computer and Vision Research Center Technical Report TR-92-13-88.
- Dhond, Umesh and J. K. Aggarwal, "Analysis of the Stereo Correspondence Process in Scenes with Narrow Occluding Objects," *Proceedings of the 11th International Conference on Pattern Recognition*, The Hague, The Netherlands, August 30 - September 3, 1992, Vol. I, pp. 470-473.
- Dhond, Umesh and J. K. Aggarwal, "Computing Stereo Correspondences in the Presence of Narrow Occluding Objects," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, 1992, pp. 758-764.
- Dhond, Umesh R. and J. K. Aggarwal, "Structure from Stereo--A Review," reprinted in *Autonomous Mobile Robots: Perception, Mapping, and Navigation*, S. Sitharama Iyengar and Alberto Elfes, Editors, IEEE Computer Society Press, Los Alamitos, California, 1991, pp. 25-46.
- Lebègue, X. and J. K. Aggarwal, "A Mobile Robot for Visual Measurements in Architectural Applications," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 195-198.
- Lebègue, X. and J. K. Aggarwal, "ROBOTEX: An Autonomous Mobile Robot for Precise Surveying," *International Conference on Intelligent Autonomous Systems (IAS-3)*, February 15-19, 1993, pp. 460-469.
- Lebègue, Xavier and J. K. Aggarwal, "Detecting 3-D Parallel Lines for Perceptual Organization," *Proceedings of the Second European Conference on Computer Vision*, May 1992, pp. 720-724.
- Lebègue, Xavier and J. K. Aggarwal, "Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot," *Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, May 1992, pp. 1778-1785.

- Lebègue, Xavier and J. K. Aggarwal, "Semantically Significant Line Segments for a Mobile Robot--ROBOTEX," *Video Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, May 1992. (Videotape presentation).
- Lebègue, Xavier, D. C. Baker and J. K. Aggarwal, "Fusion of Color and Geometric Information," in *Multisensor Fusion for Computer Vision*, J. K. Aggarwal, Editor, Springer Verlag, Berlin, 1993, pp. 213-237.
- Lebègue, Xavier and J. K. Aggarwal, "Generation of Architectural CAD Models Using a Mobile Robot," *Proc. 1994 IEEE Intl. Conf. on Robotics and Automation*, San Diego, CA, pp. 711-717.
- Lebègue, Xavier, and J. K. Aggarwal, "Automatic Creation of Architectural CAD Models," *Proc. 2nd IEEE Computer Society CAD-Based Vision Workshop*, Seven Springs, Pennsylvania, Feb. 8-10, 1994.
- Lebègue, Xavier, and J. K. Aggarwal, "Significant Line Segments for an Indoor Mobile Robot," *IEEE Trans. on Robotics and Automation*, Vol. 9, No. 6, December 1993, pp. 801-815.
- Lu, H. Q. and J. K. Aggarwal, "Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes," *Computer and Vision Research Center Technical Report RT-91-5-69*, March 1991.
- Lu, H. Q., and J. K. Aggarwal, "Applying Perceptual Organization to Detection of Man-made Objects in Non-Urban Scenes," *Pattern Recognition*, pp. 835-853, 1992.
- Lu, H. Q., and J. K. Aggarwal, "Extracting Geometric Structures and a Region of Interest for Man-made Object Detection," *Proc. 20th AIPR Workshop: Computer Vision--Meeting the Challenges*, McLean, Virginia, October 1991.
- Nair, Dinesh, and J. K. Aggarwal, "An Object Recognition System Using ART-2," *Proc. 3rd International Conference on Advances in Pattern Recognition and Digital Techniques*, December 28-31, 1993, Calcutta, India (P. C. Mahalanobis Birth Centenary Celebration).
- Nandhakumar, N. and J. K. Aggarwal, "A Phenomenological Approach to Thermal and Visual Sensor Fusion," in *Highly Redundant Sensing in Robotic Systems*, Julius Tou and Jens Balchen, Eds., NATO ASI Series, Series F: Computer and System Sciences, Vol. 58, Springer-Verlag, Berlin Heidelberg, 1990.
- Nandhakumar, N., and J. K. Aggarwal, "Multisensory Computer Vision," in *Advances in Computers*, Volume 34, Marshall C. Yovits, Editor, Academic Press, Inc., San Diego, California, 1992, pp. 59-111.
- Nandhakumar, S. Karthik, and J. K. Aggarwal, "Unified Modeling of Non-Homogeneous 3-D Objects for Thermal and Visual Image Synthesis," *Pattern Recognition*, in press.
- Rodriguez, Jeffrey J. and J. K. Aggarwal, "Matching Aerial Images to 3-D Terrain Maps," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 12, No. 12, December 1990, pp. 1138-1149.
- Sabata, Bikash and J. K. Aggarwal, "Correspondence of Surfaces in a Sequence of Range Images for Motion Estimation and Tracking," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 385-388.

- Sabata, Bikash and J. K. Aggarwal, "Estimation of Motion from a Pair of Range Images: A Review," *Computer Vision, Graphics, and Image Processing*, Vol. 54, No. 3, pp. 309-324, November 1991.
- Sabata, B., F. Arman, and J. K. Aggarwal, "Segmentation of 3D Range Images Using Pyramidal Data Structures," *Computer Vision, Graphics and Image Processing: Image Understanding*, Vol. 57, No. 3, 1993, pp. 373-387.
- Sabata, B., F. Arman, and J. K. Aggarwal, "The Convergence of Fuzzy Pyramid Algorithms," *J. Mathematical Imaging and Vision*, in press.
- Sabata, Bikash and J. K. Aggarwal, "Hypergraph Based Feature Matching in a Sequence of Range Images," *Proceedings of the 4th International Workshop on Time Varying Image Processing and Moving Object Recognition*, Florence, Italy, June 10-11, 1993.
- Sabata, Bikash and J. K. Aggarwal, "Surface Correspondence and Motion Computation from a Sequence of Range Images" *1994 IEEE International Conference on Robotics and Automation*, San Diego, CA, May 8-13, 1994, pp. 1052-1059.
- Shah, Shishir and J. K. Aggarwal, "A Simple Calibration Procedure for Fish-Eye (High Distortion) Lens Camera," *Proc. 1994 IEEE Intl. Conf. on Robotics and Automation*, San Diego, CA, pp. 3422-3427.
- Talluri, R. and J. K. Aggarwal, "Autonomous Navigation in Cluttered Outdoor Environments using Geometric Visibility Constraints," *International Conference on Intelligent Autonomous Systems (IAS-3)*, February 15-19, 1993, pp. 186-195.
- Talluri, R. and J. K. Aggarwal, "Image/Map Correspondence for Mobile Robot Self-location Using Computer Graphics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Special Issue on 3D Modeling in Image Analysis And Synthesis, Vol. 15, No. 6, June 1993, pp. 597-601.
- Talluri, R. and J. K. Aggarwal, "Position Estimation for an Autonomous Mobile Robot in an Outdoor Environment," *IEEE Trans. on Robotics and Automation*, Vol. 8, No. 5, pp. 573-584, 1992.
- Talluri, R. and J. K. Aggarwal, "Position Estimation Techniques for an Autonomous Mobile Robot - A Review," in *The Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Editors, World Scientific Publishing Co., Inc., Singapore, 1993, pp. 769-801.
- Talluri, R. and J. K. Aggarwal, "Transform Clustering for Model-Image Feature Correspondence," *Proc. IAPR Workshop on Machine Vision Applications (MVA-92)*, December 7-9, 1992, Tokyo, pp. 579-582.
- Talluri, R. J. and J. K. Aggarwal, "Mobile Robot Self-Location Using Constrained Search," *Proc. IEEE International Workshop on Intelligent Robotics and Systems*, Osaka, Japan, November 1991.
- Talluri, R. J., and J. K. Aggarwal, "Edge Visibility Regions - A New Representation of the Environment of a Mobile Robot," *Proc. IAPR Workshop on Machine Vision Applications*, Tokyo, 1990, pp. 375-380.

- Talluri, Raj, and J. K. Aggarwal, "Positional Estimation of a Mobile Robot using Edge Visibility Regions," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, pp. 714-715, June 1991.
- Wang, Y. F., and J. K. Aggarwal, "Sensor Data Fusion in Robotic Systems," *Control and Dynamic Systems: Advances in Theory and Applications, Volume 39: Advances in Robotic Systems, Part 1 of 2*, Edited by C. T. Leondes, Academic Press, Inc., pp. 435-462, 1991.
- Wang, Y. F., N. Karandikar, and J. K. Aggarwal, "Analysis of Video Image Sequences Using Point and Line Components," *Pattern Recognition*, Vol. 24, No. 11, pp. 1065-1084, 1991.

## **PARTICIPATING SCIENTIFIC PERSONNEL.**

<u>Name</u>	<u>Title</u>	<u>Degree Date</u>
Lou Hornung	Research Associate	
Fred T. May	Research Associate	
Farshid Arman	Ph.D., Computer Engineering	
Agrahara Bharatkumar	M.S. Student	
Yuh-lin Chang	Ph.D., Computer Engineering	December 1993
Vipin Chaudhary	Ph.D., Computer Engineering	December 1991
Umesh Dhond	Ph.D., Computer Engineering	August 1992
Sunil Gupta	Ph.D. Student	
Mohan Kumar	M.S. Student	
Sandeep Kumar	M.S. Student	
Xavier Lebegue	Ph.D., Computer Engineering	December 1993
Dinesh Nair	Ph.D. Student	
Jason Ng	M.S. Student	May 1992
Bikash Sabata	Ph.D., Computer Engineering	August 1993
Rajendra Talluri	Ph.D., Computer Engineering	May 1993
Mark Colaluca	Undergraduate Student	
Andrea Matthews	Undergraduate Student	
Shishir Shah	Undergraduate Student	
Phil Chu	Undergraduate Student	

## **LIST OF INVENTIONS.**

None.

## LIST OF FIGURES AND TABLES.

Figure	Title	Page
1	The Overall Structure of AIMS.	6
2	The data representation and data relationship in the framework to determine a region of interest	8
3	The HAMMER Object Recognition System	9
4	System Overview - Integration of Image Segmentation Maps Using Region and Edge Information	16
5	Perspective Projection of a Scene	17
6	The mobile robot, Robo-Tex	22